Volume Title: Output Measurement in the Service Sectors

Volume Author/Editor: Zvi Griliches, editor

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-30885-5

Volume URL: http://www.nber.org/books/gril92-1

Conference Date: May 4-5, 1990

Publication Date: January 1992

Chapter Title: Introduction to "Output Measurement in the Service Sectors"

Chapter Author: Zvi Griliches

Chapter URL: http://www.nber.org/chapters/c7230

Chapter pages in book: (p. 1 - 22)

Introduction

Zvi Griliches

> The main point is that ingenuity cannot fully or effectively compensate for lack of basic information.
>
> Kuznets (1941, 111)

The continued growth of service sectors in almost all the developing economies has fascinated and occasionally alarmed economists and other observers. The recent record in the United States is displayed in figure 1 and table 1. Overall, the relative growth of employment in services (excluding government) was rather slow until the early 1960s, the upward trend that had started in the 1920s having been interrupted by the Great Depression and the war years. The trend accelerated, however, beginning in the mid-1960s. Why should this be viewed with alarm? Table 1 provides a partial answer: productivity as measured in the national accounts has grown significantly slower in services, especially in the early postwar period, 1948–60, and in the most recent decade, 1979–89. That slowness of growth, together with the rising share of services in nominal GNP and in employment, has been viewed as a major drag on the productivity growth of the overall economy and its competitive performance.[1]

There are at least two, possibly complementary, explanations of these phenomena. The first is slower technical change in services, resulting from their

Zvi Griliches is the Paul M. Warburg Professor of Economics, Harvard University, and director of the Productivity Program at the National Bureau of Economic Research.

The author is indebted to the Bradley, Guggenheim, and Sloan foundations for financial support. He has benefited from the comments of many of the participants in this conference on an earlier draft of the introduction. The author would like to thank especially Ernst Berndt, Robert Fogel, Robert Gordon, Robert Lipsey, Marilyn Manser, and Jack Triplett for their comments on the draft as a whole.

1. Note that "commodities" do include agriculture in table 1 but not in figure 1.

Fig. 1 Relative share of nonagricultural employment: "Services" versus "commodities"

Source: Historical Statistics of the U.S., pt. 1, D127–41, and the *Economic Report of the President*, 1991, table B-43.

Note: "Services" exclude government; "commodities" exclude agriculture.

Table 1 Services versus Commodities, 1947–1989

	1947 or 1948	1960	1969	1979	1989
Share of GNP (%):*					
Current prices:					
Commodities	43.8	39.4	36.5	33.8	27.1
Services	40.1	49.9	50.7	53.3	60.7
Constant prices:					
Commodities	39.9	39.7	37.7	34.4	32.4
Services	45.5	46.2	48.0	52.5	57.7
Implicit relative price:†					
Services/Commodities	80	109	110	104	126
Relative productivity (GNP/hour):‡	108	99	102	100	88

Sources: *Mohr (chap. 1, in this vol.), *Survey of Current Business*, April 1991, 27 (for 1989), and *NIPA of the U.S., 1928–82: Statistical Tables* (for 1947 and 1948). †Ratio of shares in current and constant dollars. ‡Hours worked by industry from *Survey of Current Business*, July 1990 and the *NIPA of the U.S.*

Note: Hours and hence GNP/hour series start in 1948. The numbers for 1947–69 are not fully comparable to the 1979–89 estimates. The latter are based on the newly revised methodology described in Mohr.

intrinsically more labor intensive nature, and a potentially higher income elasticity of the demand for them (see, e.g., Baumol 1967; and Baumol, Blackman, and Wolf 1985). Only the first part of this explanation is actually worrisome. But before one accepts it as a fact, one needs to consider the second explanation, the possibility that difficulties in measuring output and prices in services may have resulted in a mismeasurement of productivity growth in these sectors, a mismeasurement that accounts for some or even much of the observed contrast with the productivity experience of commodities.

It was this latter possibility that motivated the Conference on Research in Income and Wealth to organize a conference on this range of topics, the edited proceedings of which make up this volume. An earlier conference, organized by Victor Fuchs (Fuchs 1969b), was held in Ottawa in 1967, over 20 years ago. The organizing committee of the current conference (Ernst Berndt, Timothy F. Bresnahan, Marilyn Manser, and Zvi Griliches, chair) felt that, because of the importance of the topic, newly available data, and further methodological developments, such a conference could contribute to a better understanding of the issues at hand. After a planning period of about a year, the conference was held in Charleston, South Carolina, on May 4–5, 1990.

In organizing this conference we faced the problem that services are actually a rather amorphous concept, covering a heterogeneous set of industries. Figure 2 illustrates the different historical trends of its various components: a decline in the share of employment in the transportation, communication, and public utilities industries; a continued, relatively constant rate of growth in the share of trade, finance, insurance, and real estate (FIRE) industries; and a relatively sharp growth in the share of the not-elsewhere-classified service industries, especially in business, health, and personal services.[2] Similar disparities can be seen in table 2, which records the productivity experience of the different subsectors. The average levels of labor productivity were not significantly lower in the service sectors in 1948 than in commodity production, with the possible exception of retail trade. Over time, only retail trade and other services fell significantly behind the productivity developments in commodity production. In fact, it was only relatively recently that the average of GNP per hour in all the service sectors fell below that of commodity production. But productivity growth was indeed slower in the fastest-growing subsectors: retail trade, FIRE, and services, a fact that contributed to the overall decline in the relative performance of services as a whole. Because these are also the industries where output measurement may be most difficult, the suspicion is raised that some of the observed declines could be spurious—the

2. Government is excluded from these tables and figures because our focus is on productivity measurement and only a small fraction of this sector has reasonable productivity measures. Also, the sources and determinants of growth of the government sector are somewhat different and adequate coverage would take us much beyond what could be done in this volume. Two of the chapters in this volume, Jorgenson and Fraumeni and Murray, do discuss some of the measurement issues that arise in this sector.

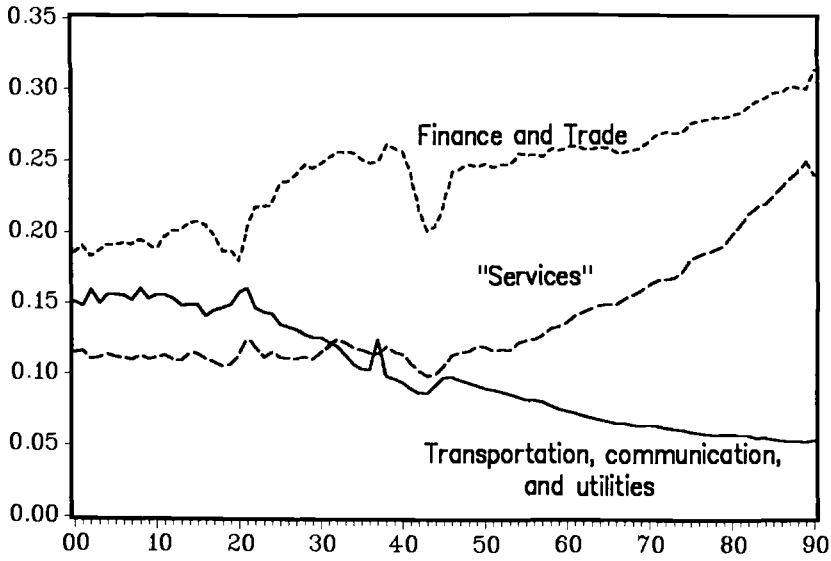

Fig. 2 Shares of total nonagricultural employment: Selected Sectors

result of our inability to observe and interpret the historical developments correctly.

To explore these issues, the committee decided to divide the conference into two parts. First, it would include a description of how services output and prices are currently measured by the major official U.S. data-collection and data-construction units: The national income accounts (GNP "originating" by industry) produced by the Bureau of Economic Analysis (BEA) in the Department of Commerce; and the consumer price index (CPI) and the productivity-by-industry series, both constructed and reported by different sections of the Bureau of Labor Statistics (BLS) in the Department of Labor.[3] And, second, it would include a discussion of alternative approaches to the measurement of output in some of these industries and studies of specific subindustries and other related topics. Before reviewing the resulting studies in some detail, it may be worthwhile to say a few more words about the slippery concept of services and why it is so problematic.

The difficulties of discussing services arise the moment one tries to define them. A standard dictionary, *Webster's Collegiate* (1946), lists as its 15th def-

3. The PPI (producer price index) is not considered explicitly in this section since its coverage of services is limited to transportation and telephone services.

Table 2 **Constant-Dollar GNP per Hour in Selected Sectors**

A. Levels (in dollars)

Sector	1948	1960	1969	1979	1989
Commodities, total	10.7	15.1	17.9	20.2	25.3
Service sectors:					
Transportation, commercial & public utility	11.6	16.4	23.0	30.2	38.2
Wholesale trade	10.0	14.5	19.1	19.7	25.6
Retail trade	8.0	9.7	11.0	12.1	14.4
FIRE	32.8	42.8	48.7	51.1	51.4
Services	10.3	11.5	13.2	14.5	15.2

B. Rates of Growth (per annum, %)

Sector	1947–60	1960–69	1969–79	1979–89	1948–89 (average)
Commodities, total	2.9	1.9	1.2	2.3	2.1
Services, total	2.1	2.2	1.0	1.0	1.6
Service – commodities	−0.8	0.3	−0.2	−1.3	−0.5
Service sectors:					
Transportation, commercial & public utility	2.9	3.8	2.8	2.4	3.0
Wholesale trade	3.0	3.1	0.3	2.6	2.3
Retail trade	1.6	1.4	1.0	1.8	1.4
FIRE	2.2	1.4	0.5	0.1	1.1
Services	0.9	1.5	0.9	0.5	1.0

inition (among 20) of the word "service" "Any result of useful labor which does not produce a tangible commodity; as, railroads, telephone companies, laundries, and physicians perform services." Note the negative definition: "not a tangible commodity." A useful definition is provided by Hill (1977) in which the user (consumer) or the user's goods are changed by the provider of the service. This definition captures the aspect of being worked on or moved either for persons, as in haircuts, physician visits (though not every psychoanalytical session results in a transformation), or airplane trips; or for goods, as in car repair, tailoring, or warehousing. One has to emphasize, however, the distinction between the production process of such activities, the question of legal ownership of the items being worked on, and the payment format—is the price paid for services rendered, a recompense for agency efforts on behalf of the principal, or a direct purchase of a commodity from a middleman? What is important in this definition is the recognition of the role of the purchaser in such a transaction, either because of her direct involvement in the activity and consequent contribution to its ultimate output or as a supplier of one of the major inputs to it. Although the quality of a commodity does not usually depend on the "quality" of its consumers (though the demand for it

may), the output of a service activity may depend on the quality and/or effort of its consumers, as in teaching and related advisory services.

Rather than discussing definitions, it may be more useful to take an operational approach and to examine what are actually called services in the national accounts and related statistical sources. The broadest definition of services corresponds to the nontangible, noncommodity notion: everything except agriculture, mining, construction, and manufacturing. This notion defines the scope of this volume but also leaves one wanting to quarrel with it from time to time. It includes transportation, communication, public utilities, wholesale and retail trade, FIRE, repair, personal, business, health, legal, and other services; and the activities of federal and local governments. It is troubled by the fact that electricity is tangible, that, although gas in bottles is a commodity, gas from a pipe is a service, and that in many cases the definition and measurement of an activity depend on rather arbitrary boundaries. If manufacturers conduct more of their trade activity out of their plant offices much of "trade" shifts back to manufacturing. Alternatively, buying prepared meals shifts output from the household sector to retail trade. Additional complications arise from the fact that governments and many nonprofit institutions do not sell their output directly, provide no relevant transaction data, and hence require a variety of more or less unsatisfactory imputation techniques.

Although services are different, they are really not so different from goods as far as the problem of measuring output is concerned. Most of the problems afflicting the measurement of commodity output affect also the measurement of services, only more so. To measure the output of any activity we need to know its total receipts and have adequate information to construct an appropriate price index for it. To measure productivity, we need in addition parallel information on the inputs used in production (total costs and prices or units used). In either case, we need to know the relevant transaction unit and deal with the problem of quality change, which arises form the underlying heterogeneity of outputs and inputs and the continuing appearance of new products, varieties, and services, and the disappearance of old ones.

Why is the problem more serious in some of the service sectors? Partly it is a data problem, but also, importantly, it is a conceptual one. Historically, much more data were collected on agricultural and manufacturing commodities and their prices than were collected on services. Censuses and annual surveys of service industries are a relatively recent development and are much less detailed in their coverage of inputs used. Many of the service industries produce intermediate products in areas with very little direct price coverage, such as computer programming, advertising, and information. The producer price index (PPI), formerly the wholesale price index, the major source of deflators for the GNP by industry series, does not collect service prices (except of air- and water-transport and telephone services). Because of this lack of data, a number of service industries series are deflated by makeshift deflators, and real output is assumed to grow proportionally to some measure of

input and to lead to no observed productivity growth by assumption. The latter is true for the whole government sector, the contribution of various nonprofit organizations, such as universities, and such difficult-to-measure sectors as banking and business services.

The conceptual problem arises because in many service sectors it is not exactly clear what is being transacted, what is the output, and what services correspond to the payments made to their providers. A whole section of this book is devoted to the discussion of how one should view and measure the output of the banking sector. Similarly, when an industrial firm keeps a legal firm on a retainer, what is the corresponding quantity of services? In several service sectors, such as business, health, and legal services, what is transacted is a delivery and exchange of information. Because of its extreme heterogeneity, it is rather difficult to price it efficiently, per bit transmitted, and therefore the resulting pricing structures are often nonlinear and not directly related to what was actually received by the consumer. This difficulty is reflected, for example, in the different pricing structures of various data services, such as Compuserve, Dialog, and Prodigy, and the associate complexity of evaluating their output.

Over all this hangs the ubiquitous issue of quality change. The problem is general and pervasive. It affects the measurement of consumer durable purchases and the measurement of output in construction. In some service industries, with good data bases and relatively homogeneous outputs, such as communication or public utilities, the measurement problems are less severe than in some of the commodity sectors. But, in general, because of the underlying heterogeneity of transactions, the difficulty of making comparisons across time and space is even greater. In many service sectors output depends on the interaction with the user and thus is more difficult to standardize. Moreover, for many commodities, even for such rapidly changing ones as personal computers and stereo equipment, one has publicly available "specification" data, which report some of the characteristics for the individual items relevant to the measurement of output and performance. The same detail is not available on the performance characteristics of doctors, lawyers, and stockbrokers. Moreover, the necessary economic-engineering research that would tell us which of the characteristics and training levels are important for their successful performance has not been done. We are thus lacking the scientific base for the desired measurement procedures.

It is best to discuss some of these issues in the more concrete context of specific industries and their special measurement problems. I shall turn, therefore, to a brief overview of the papers contained in this volume.

The volume starts with three papers on the major sources of official U.S. data in these sectors. In the first paper, Michael F. Mohr describes the recent revision in the GNP-by-industry series and its effect on the measurement of output in the service sectors. The current revision represents a significant ad-

vance on past practice, especially in its improved measurement of the intermediate inputs used in these sectors. For example, it reduces the growth in the output of health services during 1979–87 from the previously estimated 4.6 percent per year to 2.8 and reassigns the difference to the medical instruments, pharmaceuticals, and other supplier industries. But the revision is incomplete. Some major and growing subsectors, such as banking and business sectors, are still being extrapolated by input measures, eliminating productivity growth by definition. Some of the improvements come with their own problems: "revenue miles" are a reasonable measure of transportation services, but they leave open the question of quality change in these miles, a question discussed later on by Robert J. Gordon (chap. 10), and the question of the effect of various travel restrictions associated with special fares, a question considered by Paul A. Armknecht and Daniel H. Ginsberg (chap. 3). Similarly, measuring the output of brokerage services by the number of trades, treating $1,000 and $1,000,000 trades as equivalent, leaves something to be desired. Here, one could have probably constructed a reasonable index of commission rates from the Securities and Exchange Commission and other sources. One could also raise questions about the treatment of radio and television broadcasting, which under the current conventions is entirely an intermediate input, contributing only to the output of cereals and razor blades (except for public television, which is included in consumption and treated differentially). Implicitly, output in this industry is measured by the size of the audience. Thus, quality changes that expand the industry's size are reflected in output, but quality improvements that occur in the competition for audience may show up in higher rates and be "deflated" away. Because advertisers are not interested just in minimizing the cost per person reached but also in maximizing the total size of the audience reached and the effectiveness of the message, the current procedure may be problematic without even raising the issue of what consumers get out of it and how it is related to advertiser costs, if at all.

The double-deflation procedure (the subtraction of deflated intermediate purchases form deflated gross output to arrive at a real value-added concept) is itself troublesome, as is also the GNP by industry construction, which is based on a value-added measure of an industry's output and is motivated by a desire for an unduplicated measure of national output. If one is interested in productivity measurement at the industry level and has some notion of a production function as a framework for it, the subtraction of intermediate inputs from gross output is appropriate only when these inputs are used in fixed proportion to output, when the ratio of their prices to final product prices remains constant, or when changes in their prices have no effect on the relative amounts of capital and labor used in production. Neither is a very likely occurrence. If one looks at the restaurant sector, one sees that the relative price of food purchased for away-from-home consumption to food purchased for home consumption (a proxy here for the cost of intermediate input) has risen by 1.4 percent per year between 1979 and 1986 (*Survey of Current Business*,

July 1987). Similarly, in the retail food sector as a whole, the consumer price index (CPI) for food rose by 1.8 percent more per year, between 1979 and 1990, than the comparable PPI for consumer food. For productivity measurement purposes we would be much better off with explicit and separate series on gross output and intermediate inputs in constant prices. The duplication problem can be solved by using appropriate value-added weights, as was pointed out a long time ago by Domar (1961). To implement either this program or the current GNP-by-industry effort correctly requires a detailed set of productivity accounts covering the whole economy. At the minimum, as Mohr notes, we need a consistent and current set of input-output tables.[4] But the 1982 input-output table was published only very recently (*Survey of Current Business,* July 1991), with a lag of almost a decade, and is based on the old standard industrial classification. Thus, it is already obsolete.

The problem is actually deeper: it is not just the delay; it is the lack of the right underlying ingredients. To construct GNP by industry by current methods requires information either on profits and depreciation or on intermediate input purchases by industry. But, although most of the output, employment, and wage bill data are collected at the establishment level, profits and depreciation come from company-based IRS records and have to be allocated across industries based on scraps of obsolete information. At the same time, neither the census nor the annual survey of manufacturers collects a complete account of expenditures on all intermediate purchases, especially services. Moreover, the censuses of service industries often do not ask about purchases of intermediate inputs at all. Nor can the BEA in our system tell the other statistical agencies what and how to collect, to impose a consistently designed data-collection framework centered on the need for a coherent and high quality set of national income accounts. In the meantime one muddles through as best one can and the Mohr paper shows us both how much progress can be made even within the current constraints and also how far we have still to go.

The BLS program on "Productivity Measures for Selected Industries," described in the paper by Edwin R. Dean and Kent Kunze, benefits from not having to cover the whole waterfront. It concentrates on measuring productivity growth in those detailed industries where physical measures of output are available or where there is a reasonable price index for the deflation of the gross output data in current prices. But here concentration on physical measures of gross output and the lack of any information on intermediate (and capital) inputs creates problems of its own. Physical output units may also often vary widely in quality. Only electricity and gas provide us with reasonable measures of output. As noted earlier, passenger miles are a dubious unit of measure because they ignore timing and convenience considerations. Similarly problematic is the number-of-transactions approach to the measurement of banking output, as is indicated in the subsequent papers that discuss this

4. See Jorgenson (1990) for an attempt in this direction.

topic in greater detail. But measuring the output of service establishments by their deflated sales, by their "throughput," it also questionable, as can be seen from Walter Y. Oi's discussion of this topic. Looking at a fruit store and measuring its output by the number of oranges sold ignores the effort that may go into their arranging and culling. If the store stays open longer and makes itself more convenient, its measured productivity declines. If demand or supply shifts from radishes and onions to kiwis and strawberries productivity increases. The latter effect could be counteracted by deflation and weighting at the individual product level or by a CPI based on rapidly shifting weights. Neither is, unfortunately, the case in practice. Moreover, many of the CPI-based deflators, especially for electric appliances and electronic equipment but also for hotel services, may be missing quite a bit of the quality change (upgrading) occurring in all of these industries. Thus, although the spirit is willing, the execution is not always as strong as we or the BLS would desire. In his comment on Dean and Kunze's paper, W. Erwin Diewert calls attention also to the rather strange mixture of weighting schemes used to construct these measures (see, e.g., Dean and Kunze's fig. 2.4).

Nevertheless, these measures do provide us with another very useful window on reality, on what is going on in our economy. Because the BEA figures are mostly published at a much higher aggregation level, and because they are value-added, not output, measures, it is hard to make a direct comparison to the parallel BLS measures. Gordon does this comparison for the productivity measures in transportation. In table 3, I present a few of the possible comparisons for some of the service industries (see also table 2.7 in Dean and Kunze). Given the different conceptual bases (gross output vs. value added and hours vs. persons), they are often rather close. Two noteworthy differences occur in banking and automotive repairs. In both cases they arise from differences in the measurement of output. For banking, the BLS approach, although imperfect, is still superior. In automotive repairs the difference arises, presumably, from the differential treatment of intermediate inputs. Al-

Table 3 **Alternative Estimates of Productivity Growth for Selected Industries, 1979–1989**

Sector	BEA	BLS
Air transportation	0.5	1.9
Petroleum pipelines	1.1	0.4
Telephone	5.4	5.3
Banking	0.0	2.3*
Hotels	−1.1	−1.3
Automotive repair	−2.6	0.2

Note: BEA = GNP (value added in 1982 prices) per person engaged in production; BLS = Output per hour (From USDL 19-41, table 2).

*1979–88

though conceptually accounting for them is an improvement, the resulting implication of a ten-year decline in labor productivity at over 2.5 percent per year is hard to believe and is worth additional investigation. Similarly, the conclusion of both the BEA and the BLS measures, that the productivity of hotels and motels has been declining, is dubious, though it could be explained by the downward trend in occupancy rates and the increase in various ancillary services, such as concierges. Taken over all, it is good that we have two different glimpses of the same phenomena, though I keep thinking that more could be done in explaining the differences between them. Providing an explicit reconciliation could be very informative.

The paper by Paul A. Armknecht and Daniel H. Ginsburg describes the procedures currently used in constructing the CPI and some of its major services prices components. Because the CPI is the major source (together with the PPI) of the deflators used in the construction of "real" GNP, it is very important to the whole productivity-measurement enterprise. Its problems translate directly into output-measurement problems because, by and large, output is measured as deflated sales or as value added.

Armknecht and Ginsburg lay out very clearly the major issues facing the CPI: weighting, new goods and services, and quality change. Solutions to these problems depend on the availability of resources and on an agreement on what is to be measured. Both of these have been in short supply historically. The weighting problem is perhaps the most obvious one: 1960–61 base weights were used for 14 years, from 1964 until 1978; 1972–73 weights were used for the next 9 years, until 1987, by which time they were 15 years out-of-date. Current data are based on 1982–84 weights. Actually the problem is not as bad as it sounds because as of 1978 the CPI started adjusting some of the internal weights and shifting the product mix to be priced.[5] Probability sampling procedures introduced in 1978 in principle bring new items into the index within a five-year cycle. But such a two-to-three-year average lag is still too long in a world of rapidly changing products with most of the price declines occurring in the first few years after their introduction. Thus, for example, personal computers did not enter the index until 1987, and many of the new models do not live long enough to be caught in such a sampling cycle.

A more general problem arises from the standard "linking" procedure for new goods in both the CPI and PPI: goods are defined too finely, and hence the gains from their appearance are "linked out" from the index. For example, video rentals are a lower-cost alternative to movies but the transition to them does not lower the entertainment price index as a whole. Nor does the appearance of generic drugs show up as a decline in pharmaceutical prices, because the generic version of a branded item is treated as a separate commodity.

5. Overall weighting is probably less a problem (see Manser and McDonald 1988) for the CPI than for the GNP implicit deflator. The lag in the introduction of new products and the treatment of new outlets (Reinsdorf 1990) may be empirically more important.

Armknecht and Ginsburg describe a very interesting attempt to use hedonic regression adjustments in computing the price of air travel, which could also be used to improve the parallel BEA and BLS estimates. They also discuss the very difficult issue of measuring the price of health services and health insurance. Here we come up squarely against the question, What is to be priced? What is the service of a physician—a consultation, a procedure, or a cure? How do we adjust for quality change if what the physician does or suggests is more effective than it used to be? Are insurance rate increases, caused by a rise in morbidity or by the increased use of new procedures, a rise in price or in quantity?

Armknecht and Ginsburg discuss the dilemmas in this area and conclude that an increase in the utilization of the health-care system because of the appearance of a new disease should not affect the relevant price indexes even though health-insurance rates may rise. This is right, but it leads to the paradox of an increase in "real" GNP that is likely to be misinterpreted as a rise in the standard of living. Ideally, one would like to distinguish between the price index of living and the cost index of living, of keeping the consumer at some fixed utility level. The cost of living may change because prices have changed, or it may change because the physical and social environment has changed. Thus, one may wish to exclude taxes from the definition of the price index and above normal expenditures on heating or health from the dual definition of the level of living. The use of the concept of a household production function (Becker 1965), extended to include a separate disturbance to the technology of consumption, would break the identity between expenditures in constant prices and the associated indirect utility index. This breakage would lead us toward a redefinition of the GNP concept to allow for capital gains and losses resulting from natural disasters, epidemics, and the depletion or restoration of various natural resources. We are still very far from having the data bases necessary for doing this right, but it is worthwhile to try to work out the conceptual problems of extending the national accounts in these directions.[6]

In the meantime, however, "simpler" quality-change problems also require attention. Statistical agencies have been quite reluctant to move in the direction of pricing a "cure" or a "disease episode," partly for conceptual reasons (see, e.g., the old exchange between Gilbert 1962 and Griliches 1962), but mostly because of the difficulty in collecting the relevant data. Here, the increased movement toward payments by diagnostically related grouping (DRG) may help. But ignoring the problem may also result in serious biases. For example, the new laser-based gall-bladder procedure has reduced significantly the total cost of treating such episodes. Because it is a different procedure, it does not show up as a decline in the "real price" of health services. But, assuming that it does not lead to an increase in the total number of oper-

6. These are not new ideas. See Kuznets (1941) and Nordhaus and Tobin (1972).

ations, the effect of such a substitution is a *decline* in the output of the health-services sector and the resulting reduction in resources used does not show up as a productivity increase.

I believe that we can and should do a better job of tracking such changes and incorporating them into our measurement procedures. It should be noted, however, that, despite the implicit complaint above, the CPI is probably the best of all the statistical series produced by the U.S. government, in the extent of its attention to this range of problems and the effort it puts into guarding and improving the quality of the primary data that it collects.

The next set of papers discusses the productivity experience of specific industries while at the same time straining against the conventional boundaries of national income accounting and trying to go beyond them, often quite far, in the quest for alternative measurement procedures. Walter Y. Oi reviews the measurement problems in the retail trade sector and the available empirical evidence on them. He emphasizes the importance of inventories and inventory services provided by this sector to consumers with the resultant implication of significant economies of scale (what he calls the economies of massed reserves). He also notes that the shifting boundaries of activity between manufacturers, wholesalers, retailers, and consumers make conventional productivity measurement both difficult and often misleading. Thus, for example, current measures of double-deflated output in the retail food industry underestimate its growth and the associated productivity increases because the CPI links out the price decline that occurred with the introduction and spread of chain stores by treating them as a separate commodity. At the same time, the BLS measure of the productivity of gasoline stations overestimates their productivity growth by excluding the growing use of consumer input in self-service stations. A related measurement problem affects also some of Oi's own cross-sectional comparisons: looking at sales per employee or similar measures across different-size stores underestimates the true extent of economies of scale because it does not take into account the lower price levels in the larger stores.

It would be useful to have price measurement not only across time, as in the CPI or PPI, but also across space and type of outlet, holding "true" service levels constant. This task is very difficult, similar in magnitude to that undertaken by the United Nations International Comparison Project reported on by Alan Heston and Robert Summers in their paper in this volume.[7] But without something along such lines we are unlikely to make real progress on productivity measurement in this area. To escape some of these difficulties Oi suggests that we may have to give up the quest for measuring productivity separately at each transaction level and concentrate instead on looking at the

7. Experimental work along these lines is currently being pursued by the BLS in its Interarea Price Program.

"final" productivity of the product delivery process as a whole: cake on a plate or gasoline in the tank. Gordon pushes this idea even further below and suggests miles driven rather than gallons of gasoline in the tank as the final measure.

Timothy F. Bresnahan, Paul Milgrom, and Jonathan Paul take up the very difficult question of the output of the stock market and the associated brokerage and finance and information services. Following the earlier lead of Samuelson (1957), they emphasize that much of the activity in the stock market is devoted to anticipating next week's, next hour's, and even next minute's price, resulting in significant wealth transfers and rent dissipation. They show that the contribution of this activity to "real" productivity, either via its use in managerial compensation schemes or via its influence on investment decisions, is likely to be very small because the signal provided by the market is about the average value of the firm and that is only very loosely related to the marginal contribution of a particular managerial action or new investment project. On the other hand, because acquiring new information may result in a transfer of rents, this may induce excessive investments in such information gathering activities (see Hirshleifer 1971, for a similar argument in a different context). Given the great decline in the cost of communication and computing, it is not surprising that both trading on and employment in the stock markets has expanded greatly. The authors develop a model of such activity that indicates that its growth may be a mixed blessing and that the number of transactions on the market (the measure used by the BEA to move its real output series) may not be a good indicator of its social product. Their point could be expanded to attack also the puzzle of why the contribution of computers to productivity growth is not visible in the usual measures (see Baily and Gordon 1988; and Donald Siegel and Zvi Griliches, chap. 11, in this volume) and to inquire into the productivity consequences of the recent growth in telemarketing. Related issues of what is the output in the insurance sector and how it is to be measured were also discussed in a paper by A. Hornstein and E. C. Prescott (1991), not included in this volume.

Conceptually simpler but empirically still very intractable is output measurement in the banking industry. A separate session with two papers and three discussants was devoted to this topic. Currently, even the nominal value of banking output is in dispute. For obscure reasons, discussed in the comment by Jack E. Triplett, interest received for loans made is not counted and instead an imputation is made that is substantially greater than the interest "foregone" on demand deposits, to reflect the flow of consumer services provided by banks. Over time, output is extrapolated by the BEA by employment growth and by the BLS Industry Productivity Program as a weighted average of checks cleared and loans made, neither of which seems fully to capture what banks are all about.

To estimate the output of banks, Dennis J. Fixler and Kimberly D. Zieschang use a translog distance function (essentially a joint production func-

tion) that relates eight different financial assets, such as different types of loans and deposits, to three "conventional" inputs: labor, physical capital, and materials. The resulting estimates produce an opportunity cost of funds series and are used to construct flows of financial services associated with each of the specific asset types. They imply an output index that rises by 8.8 percent per year during 1984–88, much faster than the parallel estimate by BLS of 3.6 percent per year and the miserly 0.7 percent per year estimated by the BEA. This index may still be an underestimate because it does not take into account important quality changes that have occurred in this industry from the point of view of its consumers, especially the spread of and the improved convenience in the use of automated teller machines (ATMs) and other electronic funds transactions.

Another view of the productivity experience of banks is taken by Allen N. Berger and David B. Humphrey who use a cost function with a thick-frontier approach to measure their performance. They report an actual decline in total factor productivity in banking in recent years and attribute it to deregulation and the subsequent dissipation of rents. Because output is measured from the cost side of banks rather than from the utility of the services rendered by consumers, their output measure (or anybody else's) does not capture the increased supply of services arising from the resulting competition for depositors. Berger and Humphrey are aware of this distinction. They show that some of their decline in output was passed on to consumers in the form of an increase in interests payments on deposits and also note that costs associated with ATMs resulted in increases in services that were enjoyed by depositors but could not be incorporated in their output measure.

The Berger and Humphrey finding of widespread inefficiencies is challenged by Frank C. Wykoff in his comment. He interprets these inefficiencies as possibly arising from differences in location and from product differentiation, aspects of which are not fully taken into account in the estimated cost function. On the other hand, the recent experience of the banking sector makes their findings more credible. Jack E. Triplett, in his comment, suggests a hedonic approach to output measurement in this industry, with data on both loan charges and service charges and the "free" services associated with different types of deposit accounts to be used in estimating price indexes for bank services. Such price indexes could then be used to deflate the nominal receipts of banks and produce a more appropriate quantity index of bank services.

Education and health services are probably the most difficult sectors for output measurement. Even though we tried, we did not succeed in including a paper on the measurement of health-services output in this conference. But on education we do have the pioneering work of Dale W. Jorgenson and Barbara M. Fraumeni. In a series of papers (see also Jorgenson and Fraumeni 1989 and 1992) they have suggested a new measurement procedure and have implemented it on U.S. data. In essence, their approach defines the output of the educational system in a particular year as the net addition to human capital

that occurs as the result of the various student bodies completing an additional school year. The value of this addition is derived from current wage-age-schooling relationships that are projected into the future and then discounted back into the present. The procedure adopted makes a number of controversial assumptions: it accepts differences in the existing wage structure as reflecting primarily differences human capital produced by the educational system and steps over issues of discrimination and selectivity by ability and by socioeconomic status (on the latter, see the earlier discussions in Denison 1964; Griliches 1970; and Willis 1986). It also assumes that leisure time is to be valued at the same wage rate as working time, an assumption that is questioned in Michael Rothschild's comment on this paper. The use of current wages as fully reflecting the correct expectations about the future could also be questioned. It implies a set of capital gains-and-losses terms in the associated wealth accounts (see Jorgenson and Fraumeni 1989). The resulting estimates are "gross" in the sense that they do not allow yet for the input of student and teachers time, capital, and possibly most importantly, family time (including child rearing) used in "producing" some of this output. The former, excluding the child-rearing component is included in their subsequent paper for the Uppsala Conference (Jorgenson and Fraumeni 1992). In spite of these reservations, their numbers do draw attention to the fact that investment in human capital represents the major investment activity of this and other economies—something that is often overlooked in the various policy debates.

The paper by Swati Mukerjee and Anne Dryden Witte on the day-care industry could have, in principle, benefited from following the Jorgenson and Fraumeni lead. The effective output of that industry is some combination of parental hours "relieved" and the present value of the increase in the human capital of the children as the result of the various training and educational activities pursued there. The data are not available, however, to pursue such a path in this industry. Instead, Mukerjee and Witte use a cost-function framework and child hours (adjusted for age differences) as their primary measure of output and concentrate on developing a "quality of child care" measure based on staff hours per child. They show that this measure of quality affects costs significantly and also that it appears to have declined nationally as child-care institutions were expanding at a fast rate. Although questions can be raised about the exogeneity of staff-pupil ratios in such a cost-function context, their results do suggest that the growth in the output of this industry may be overestimated when the concomitant decline in its quality is ignored.

Robert J. Gordon's paper brings us back to more traditional ground. Reviewing the construction of output data in the transportation sector in some detail, it focuses especially on the railroad and air-transport industries and on the trucking industry. The paper is organized around the development of multifactor productivity indexes for these industries, incorporating his newly developed capital estimates (Gordon 1990). Especially noteworthy are his findings that deregulation did not reduce the quality of air transport significantly

and that incorporating public infrastructure capital into the productivity accounts does not change the productivity growth story by much. Going beyond the current measurement conventions in including fuel efficiency gains in his measurement of the changing quality of capital equipment, he argues that the ability of an airplane to generate net revenue is the appropriate starting point for measuring the relevant aircraft price indexes. He also computes an estimate of consumer time saved as a result of the improvement in flight speeds. The latter calculation yields a very large number, but one that originates largely in the airplane manufacturing industry rather than in the air-transport industry and was realized primarily in the previous several decades. More recently (1978–87), multifactor productivity in the air transportation industry is estimated by Gordon to have grown at only 1.3 percent per year as compared to 5.5 percent in the first postwar decade. Compared to other service industries, however, the productivity performance of the transportation sector has held up reasonably well.

The remaining four papers are more heterogeneous. The paper by Donald Siegel and Zvi Griliches started out from the premise that, if the output of services is not measured correctly, its contribution should show up in the measured productivity of those industries that use these services. Because the census of manufactures reports on the purchases of some services, primarily communication and repair services, the idea was to correlate this information with multifactor productivity growth for the same four-digit SIC-level industries in manufacturing. In trying to implement it they ran into serious difficulties as a result of sampling problems and sample change problems in the underlying annual data. Their paper digresses, therefore, to consider other aspects of these data, including the growing use of foreign inputs and the associated mismeasurement of their prices, and the effect of the increased use of computers in manufacturing. Their main finding is a negative one: the recent recovery of productivity growth in manufacturing cannot be attributed to increases in purchased services, foreign outsourcing, or a decline in the quality of the data. They do find, however, a positive correlation between productivity growth in different four-digit-level industries and the intensity of their investment in computers.

Elizabeth Kremp and Jacques Mairesse use the French survey of service industries to examine cross-sectional differences in productivity at the firm level for selected service industries in France. They construct a large and detailed panel-data set for their study and analyze the experience of over 2,300 French service firms. What strikes one there is the extreme heterogeneity in the experiences of these firms. For example, even though legal services firms have both a higher average labor productivity level and a much higher productivity growth rate (1984–87) than personnel supply firms (temporary-employment agencies), the firm distributions of value added per worker overlap greatly in these two industries in levels and almost completely in rates of

growth. They find significant industry as well as strong locational differentials (price-level differences?) in productivity levels but only very little in productivity growth rates. Nor is productivity related to firm size. Not having more detail on the output characteristics of these firms, it is hard to make much progress on the output-measurement issue (in the absence of decent deflators at the detailed industry and firm level). The availability of such survey data does, however, open up the possibility of studying other interesting questions (such as exit and entry behavior) about the functioning of service firms in a modern economy (see Kremp and Mairesse 1992, for a further analysis of these data).

The paper by Alan Heston and Robert Summers reports on only one aspect of a much larger enterprise: the United Nations International Comparison Project. It describes the problems that arise in making international comparisons of service prices and the various solutions adopted in this work. It then presents estimates of nominal and "real" shares of services in consumption and GDP in 1980 for 60 countries, where "real" means that the various service flows are valued at a common average set of international prices rather than in varying domestic prices. Their main finding is that the real share of services rises very little with income but that nominal service shares rise primarily because of higher relative service prices in higher-income countries. This finding is consistent with Baumol's hypothesis that the relative labor intensiveness of services raises their price as income and real wages go up. Combined with the relatively low-price elasticities estimated by Heston and Summers, it will result in an ever-growing nominal share of these industries in consumption and GDP. But the higher nominal shares do not imply a higher real consumption of such services, only a higher expenditure on them.

The last paper in this volume touches on a very important topic: the productivity of the public sector. In it Richard Murray summarizes the results of a large-scale attempt in Sweden to measure the output and productivity of various public bodies and enterprises, such as police, weather forecasting, education, and hospitals. Although the measurement problems are horrendous, the Swedish study tried to develop outcome, rather than input or throughput measures. It compared the resulting output measures to total inputs used, including capital, and produced, in effect, a multifactor productivity index for the whole public sector. The findings of the Swedish study, which may surprise some, indicate a continued and pervasive productivity decline in most public-sector activities during the periods surveyed (primarily 1960–80). Murray discusses extensively possible biases that could arise from a neglect of positive qualitative developments but concludes, after examining a variety of indicators, that they are unlikely to have done so enough to overturn the original findings. For example, neither life-expectancy nor morbidity statistics showed much improvement in the period examined to negate the finding of productivity decreases in health services. Similarly, the observed crime and recidivism rates do not conflict with the finding of a productivity decline in

the justice and police system. A possible interpretation of these findings is that various demand shifts have drawn in more resources into areas that have encountered sharply diminishing returns. In health services, new technological possibilities have expanded the scope of possible medical interventions with only marginal improvements in the overall health status of the population. At the same time, societal-environmental changes have increased criminal activities and reduced the effectiveness of the existing legal and police systems.

What is surprising is that the same trends do not show up in the U.S. statistics. The BLS productivity measurement program for selected government services (reported in BLS Bulletin no. 2349 and reviewed in Kendrick 1991) yields an overall 1.4 percent per year improvement in labor productivity in the covered portion of federal government activities for the 1967–88 period. In part the difference arises from the fact that the Swedish estimates are for total factor productivity rather than just labor productivity, but mostly, I believe, because the U.S. numbers are based more on "activity" rather than on "outcome" measures. For example, in the case of the legal system the distinction is clear: the U.S. measures are based primarily on cases handled; the Swedish study measures cases solved. One may also question the veracity of the estimated 1.7 percent per-year improvement in the productivity of federal education and training activities or the 1.2 percent per year improvement (from 1967 through 1988) in the productivity of the U.S. postal service. Such measures need to be improved to take consumer and producer satisfaction more into account.

The papers collected in this volume illustrate the great heterogeneity that hides behind the general label "services" and also the difficulty of getting a good handle on what is actually happening there. Contrasting them to the papers included in the Fuchs (1969) volume, one does find significant progress in the official data series and a wider understanding of the problem and difficulties involved in the measurement of any economic activity. Progress is reflected in the work reported here on transportation, education, and other service sectors. But several areas remain as difficult today as they were then. The problem of measurement in the health-care industry (not covered in this volume) are not much closer to solution now. Nor is the debate about the measurement of the output of banks, outlined by Gorman in the Fuchs volume settled yet. Neither do we have the tools, currently, to resolve the measurement of output issues in retail trade and related service sectors, issues that were already discussed then by Barzel, Schwartzman, and others, and now restated and expanded by Oi in this volume. New issues are raised in this volume by Bresnahan, Milgrom, and Paul, by Oi, by Murray, and by others, but they all founder on the lack of relevant data about the uses of consumer time and household and firm activity outside the conventional market sphere.

The exclusion of the household sector from the national income accounts

was originally an expedient compromise. The time may have come, however, to move toward its inclusion in the next revision of the accounts. This inclusion will require an extension of the current population survey and/or the consumer expenditure survey toward the collection of much more data on time use and household activity. But without some new data of this sort we will not be able to evaluate productivity trends in many important service industries where the primary effect of technical change has been not in terms of the items themselves, but in what they accomplish when used in the household sector and how they substitute for consumer time and other purchased inputs. The growth of the entertainment industry, previously alluded to, the substitutions arising from movie video rentals, the rise of the fast food and take-out industries, and the introduction of microwave ovens have all reduced significantly the time used in household production and have actually raised the average quality of the final product. Such contributions of services to household productivity, to improvements in health, and to increases in human capital cannot really be measured without new and more extensive data. Although new data-collection efforts inevitably turn out to be incomplete and imperfect, they are still worth pursuing because the alternative of not knowing, of giving up, is a defeat—a defeat not only in terms of not having measured something that we would like to know but also because our understanding of what we do measure is heavily affected by where we draw such boundaries and how firm they stay put over time.[8]

In the meantime, the actual productivity situation may not be as bad as some of the crude numbers indicate. In some sectors, such as communication, where we have good data, productivity is growing at a satisfactory rate. In others where our measurement efforts are still in their infancy, we should not overinterpret the numbers. Overall, I am more sanguine about the underlying productivity growth possibilities than some of the other commentators. It is true that baby-sitting may not be a beneficiary of productivity-improving developments, but the development of intercoms and the day-care industry have provided alternative and often more "productive" ways of satisfying some of the same wants. Similarly, taking Baumol's favorite example, it still takes the same four people the same time to play one of Beethoven's quartets, but their productivity, in terms of audiences reached has improved greatly, especially when recordings, radio, and television are taken into account. This fact is even true of live performances as the result of larger halls, alternative venues such as stadia, and lower real transport costs. The economies of scale inherent in modern communication media have created the phenomenon of superstars (Rosen 1981). In a real sense, Pavarotti is much more productive today than Caruso was in his own time (and also better paid). This trend may not be true, of course, of all services. In some areas, such as health and the criminal justice system, we may be facing sharply diminishing returns in spite of the many

8. I am indebted to Robert Fogel for some of the ideas in this paragraph.

technological improvements that may have affected them. But unless we improve our measurements in this area, both in terms of the availability of basic statistics and improvements in the conceptual frameworks for their interpretation, we will never know. It is the hope of this volume to have taken a small step in this direction.

References

Note: This list contains also a number of relevant references not quoted explicitly in the introduction.

Baily, M. N., and R. J. Gordon. 1988. The Productivity Slowdown, Measurement Issues, and the Explosion of Computer Power. *Brookings Papers on Economic Activity* 2:347–431.

Baumol, W. J. 1967. Macroeconomics of Unbalanced Growth: The Anatomy of Urban Crisis. *American Economic Review* 57(3): 415–26.

Baumol, W. J., S. A. B. Blackman, and E. N. Wolf. 1985. Unbalanced Growth Revisited: Asymptotic Stagnancy and New Evidence. *American Economic Review* 75(4): 806–17.

Becker, G. 1965. A Theory of Allocation of Time. *Economic Journal* 75(294): 493–517.

Bureau of Labor Statistics. U.S. Department of Commerce. 1983. *Measuring Productivity in State and Local Government.* BLS Bulletin no. 2166. Washington, D.C.: Government Printing Office.

———. 1990. *Productivity Measures for Selected Industries and Government Services.* BLS Bulletin no. 2349. Washington, D.C.: Government Printing Office, February.

Denison, E. F. 1964. Measuring the Contribution of Education. In *The Residual Factor and Economic Growth,* 13–55, 77–102. Paris: OECD.

Domar, E. 1961. On the Measurement of Technological Change. *Economic Journal* 71:709–29.

Fuchs, V. R. 1969a. *The Service Economy.* New York: Columbia Univ. Press.

———, ed. 1969b. *Production and Productivity in the Service Industries.* NBER Studies in Income and Wealth, vol. 34. New York: Columbia Univ. Press.

Gilbert, M. 1962. Quality Change and Index Numbers: A Reply. *Monthly Labor Review* 85, no. 5 (May): 544–45.

Gordon, R. J. 1990. *The Measurement of Durable Goods Prices.* Chicago: Univ. of Chicago Press.

Griliches, Z. 1962. Quality Change and Index Numbers: A Critique. *Monthly Labor Review* 85, no. 5 (May): 532–44.

———. 1970. Notes on the Roles of Education in Production Functions and Growth Accounting. In *Education, Income, and Human Capital,* ed. W. L. Hansen, 71–114. NBER Studies in Income and Wealth, vol. 35. New York: Columbia Univ. Press.

Grubel, H. G., and M. A. Walker. 1989. *Service Industry Growth.* Vancouver, B.C.: Fraser Institute.

Hill, T. P. 1977. On Goods and Services. *Review of Income and Wealth* 123(4): 315–38.

Hirshleifer, J. 1971. The Private and Social Value of Information and the Reward to Inventive Activity. *American Economic Review* 51(3): 561–74.

Hornstein, A., and E. C. Prescott. 1991. Insurance Contracts as Commodities: A Note. *Review of Economic Studies* 58:917–28.

Inman, R., ed. 1985. *Managing the Service Economy.* New York: Cambridge Univ. Press.

Jorgenson, D. W. 1990. Productivity and Economic Growth. In *Fifty Years of Economic Measurement,* ed. E. R. Berndt and J. E. Triplett, 19–118. NBER Studies in Income and Wealth, vol. 54. Chicago: Univ. of Chicago Press.

Jorgenson, D. W., and B. Fraumeni. 1989. The Accumulation of Human and Nonhuman Capital, 1948–84. In *The Measurement of Saving, Investment, and Wealth,* ed. R. E. Lipsey and H. S. Tice, 227–82. NBER Studies in Income and Wealth, vol. 52. Chicago: Univ. of Chicago Press.

———. 1992. Investment in Education and U.S. Economic Growth. *Scandinavian Journal of Economics.* Forthcoming.

Kendrick, J. W. 1991. Appraising the U.S. Output and Productivity Estimates for Government: Where Do We Go from Here? *Review of Income and Wealth* 37(2): 149–58.

Kremp, E., and J. Mairesse. 1992. A Look at Productivity Level in French Service Industries, 1984–1988. *Journal of Productivity Analysis.* Forthcoming.

Kuznets, S. 1941. *National Income with Composition, 1919–38.* New York: NBER.

Manser, M., and R. McDonald. 1988. An Analysis of Substitution Bias in Measuring Inflation, 1959–85. *Econometrica* 56(4): 909–30.

Nordhaus, W. D., and J. Tobin. 1972. Is Growth Obsolete? *Economic Research: Retrospect and Prospect.* Vol. 5, *Economic Growth.* 50th Anniversary Colloquium. New York: NBER.

Ofer, G. 1973. *The Service Sector in Soviet Economic Growth.* Cambridge, Mass.: Harvard Univ. Press.

Reinsdorf, M. 1990. The Effect of Outlet Price Differentials on the Consumer Price Index. Paper presented at the NBER Conference on Price Measurements and Their Uses. Washington, D.C., March 22–23.

Rosen, S. 1981. The Economics of Superstars. *American Economic Review* 71(5): 848–58.

Samuelson, P. A. 1957. Intertemporal Price Equilibrium: A Prologue to the Theory of Speculation. *Weltwirtshaftliches Archiv* 181–219.

Searle, A. D., and C. A. Waite. 1980. Current Efforts to Measure Productivity in the Public Sector: How Adequate for the National Accounts? In *New Developments in Productivity Measurement and Analysis,* ed. J. W. Kendrick and B. N. Vaccara, 333–56. NBER Studies in Income and Wealth, vol. 44. Chicago: Univ. of Chicago Press.

Willis, R. J. 1986. Wage determinants: A Survey and Reinterpretation of Human Capital Earnings Functions. In *Handbook of Labor Economics,* vol. 1, ed. O. Ashenfelter and R. Layard, 525–602. Amsterdam: Elsevier-North Holland.

Wolfson, M. C. 1991. A System of Health Statistics: Toward a New Conceptual Framework for Integrating Health Data. *The Review of Income and Wealth* 37(1): 63–80.